

Customer churn analysis

How decrease the customer churn rate

By: **Johny Vallejo** - *Data Scientist, Mi Águila*

Agenda

- Introduction
- EDA
- Modelling
- Predict
- Recommendations

Introduction

Company

Telecommunication company EagleSouth, a US-based Telecom.

EagleSouth has records since 72 months ago of them users.

Context

The company is concerned about customer retention.
We've a database:

- 6,143 **users**
- 19 **variables**:
 - 3 numeric
 - 16 categorical

Problem statement

Design a strategy to both understand the problem and implement new, cost-effective solutions.

EDA

Exploratory Data Analysis

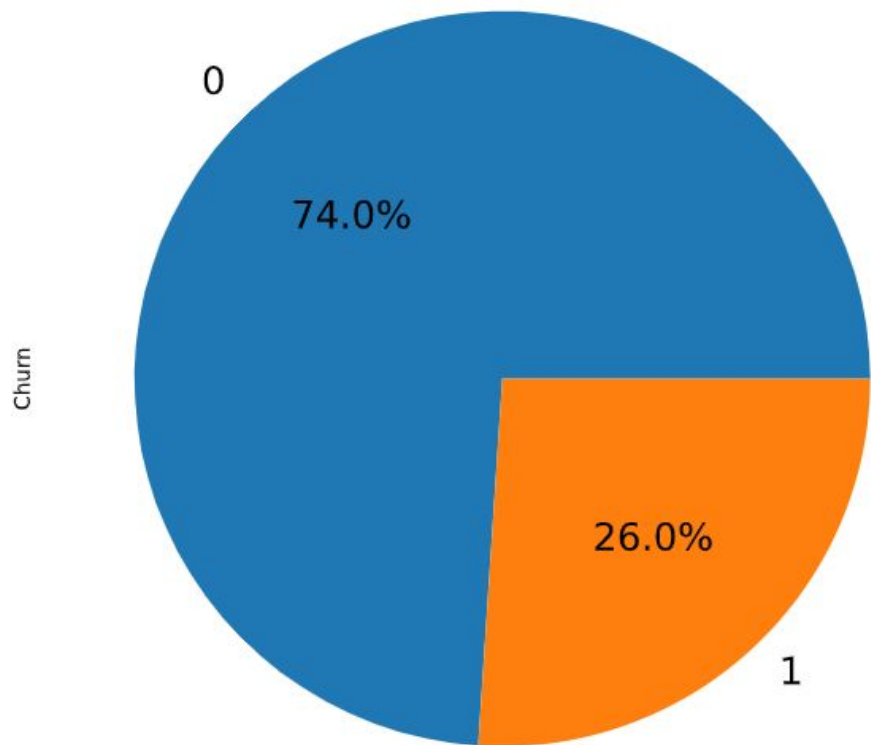
1	gender	6143	non-null	object
2	SeniorCitizen	6143	non-null	int64
3	Partner	6143	non-null	object
4	Dependents	6143	non-null	object
5	tenure	6143	non-null	int64
6	PhoneService	6143	non-null	object
7	Multiplelines	6143	non-null	object
8	InternetService	6143	non-null	object
9	OnlineSecurity	6143	non-null	object
10	OnlineBackup	6143	non-null	object
11	DeviceProtection	6143	non-null	object
12	TechSupport	6143	non-null	object
13	StreamingTV	6143	non-null	object
14	StreamingMovies	6143	non-null	object
15	Contract	6143	non-null	object
16	PaperlessBilling	6143	non-null	object
17	PaymentMethod	6143	non-null	object
18	MonthlyCharges	6143	non-null	float64
19	TotalCharges	6143	non-null	object
20	Churn	6143	non-null	object

Drop records of new users

customerID	gender	SeniorCitizen	Partner	Dependents	tenure
A00307	Female	0	Yes	Yes	0
A00545	Female	0	Yes	Yes	0
A00936	Male	0	No	Yes	0
A02162	Male	0	Yes	Yes	0
A03137	Female	0	Yes	Yes	0
A03367	Male	0	Yes	Yes	0
A03555	Male	0	Yes	Yes	0
A04557	Female	0	Yes	Yes	0
A04862	Male	0	No	Yes	0
A04971	Female	0	Yes	Yes	0
A05316	Male	0	Yes	Yes	0

There're **11 new users** that was dropped from the dataset for avoid skewed the model.

Churn percent in the dataset

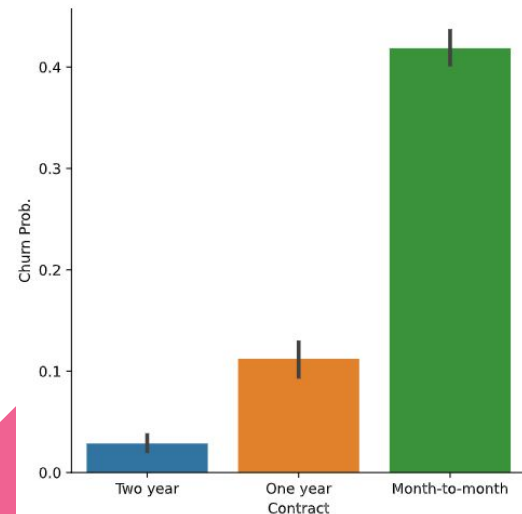
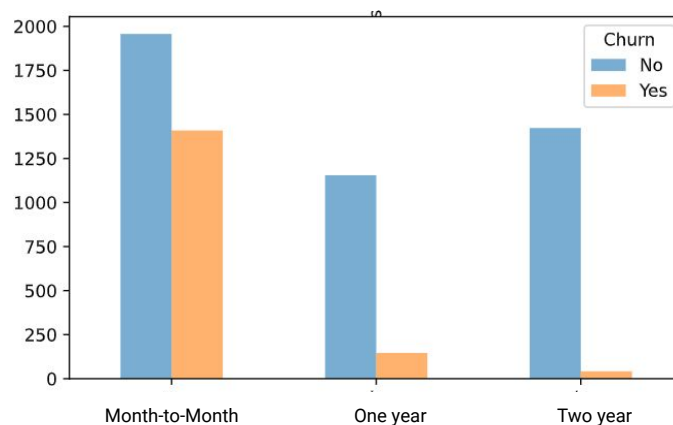
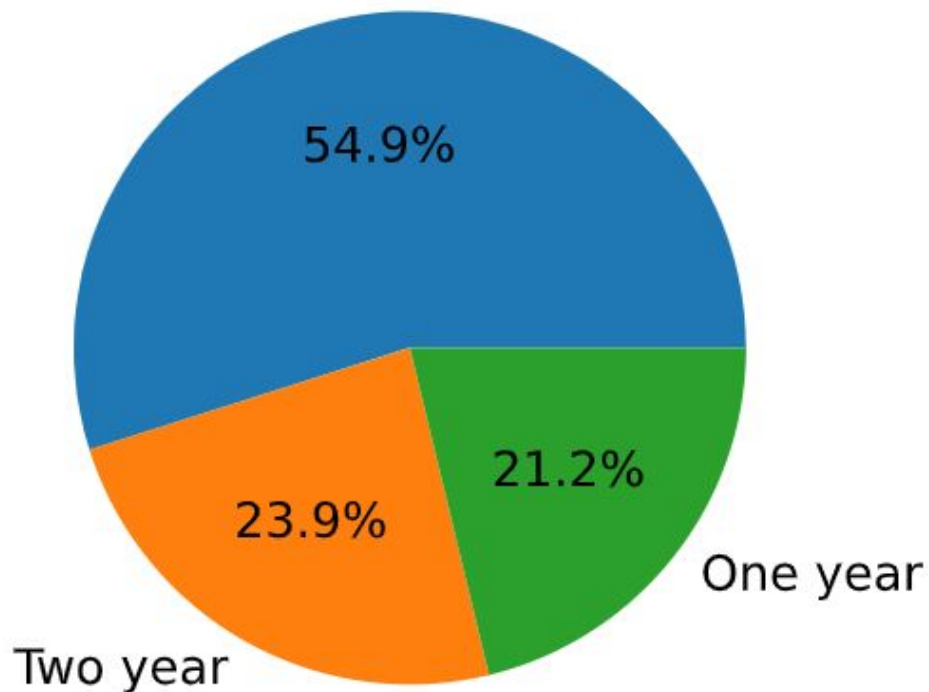


Our scope has 74% instance of **no churn (0)**. Maybe the ML-models will be skewed for insufficient instance of **churn (1)**.

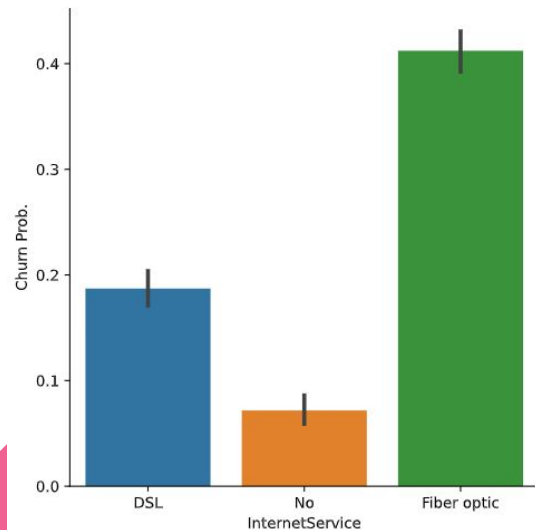
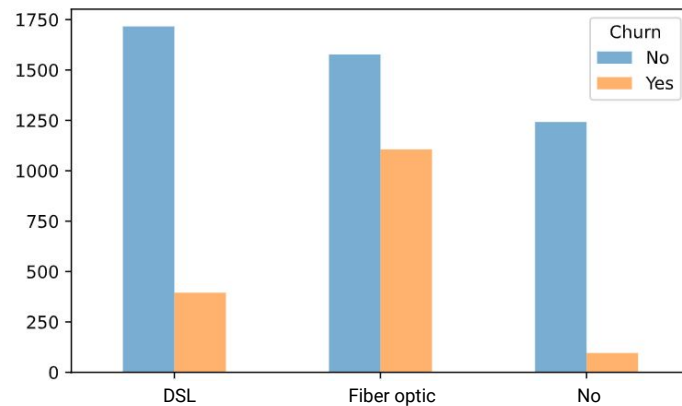
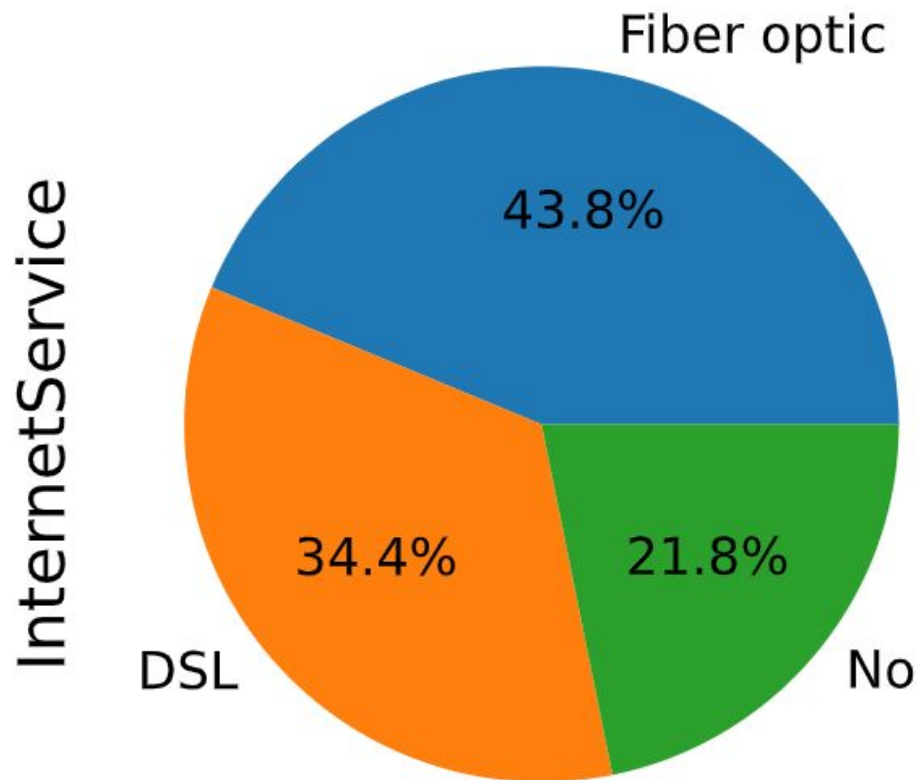
Categorical variables (1/3)

Month-to-month

Contract

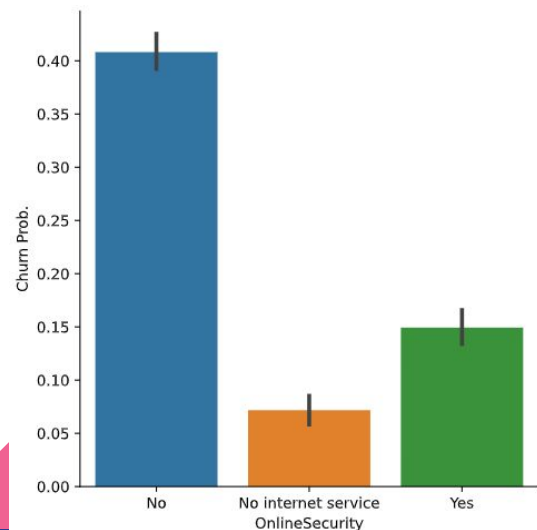
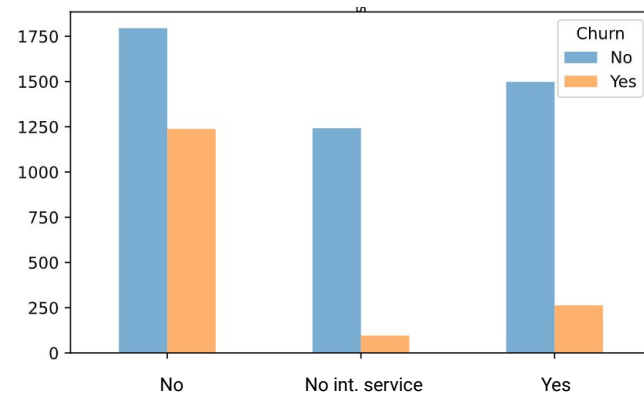
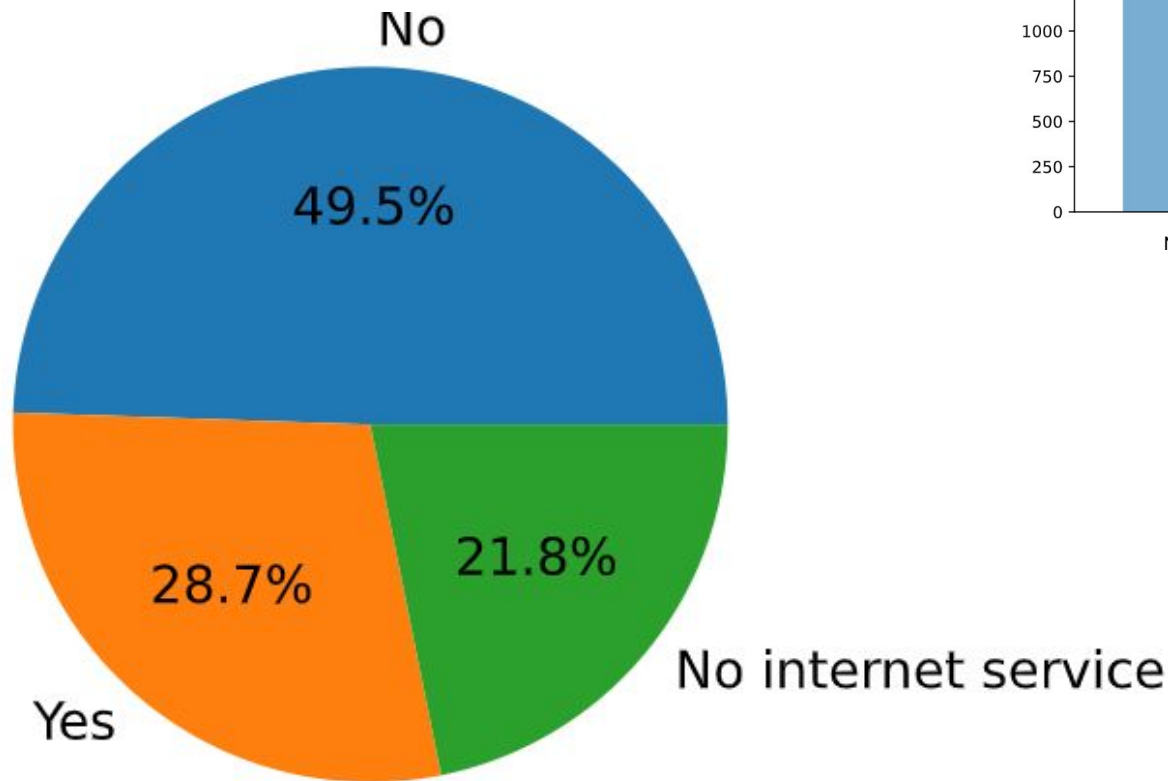


Categorical variables (2/3)

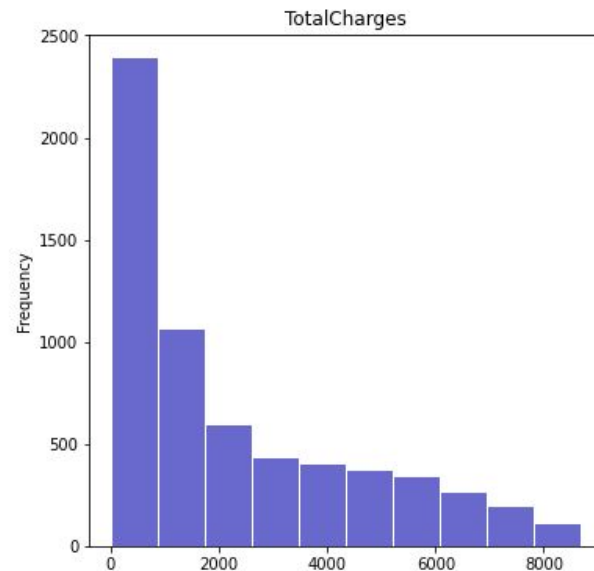
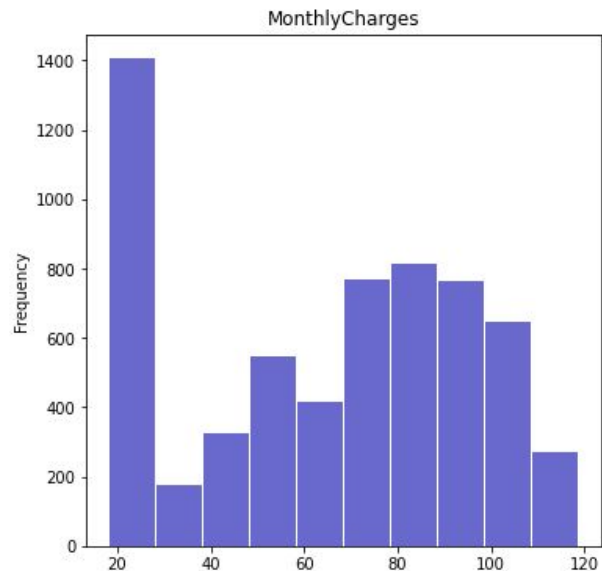
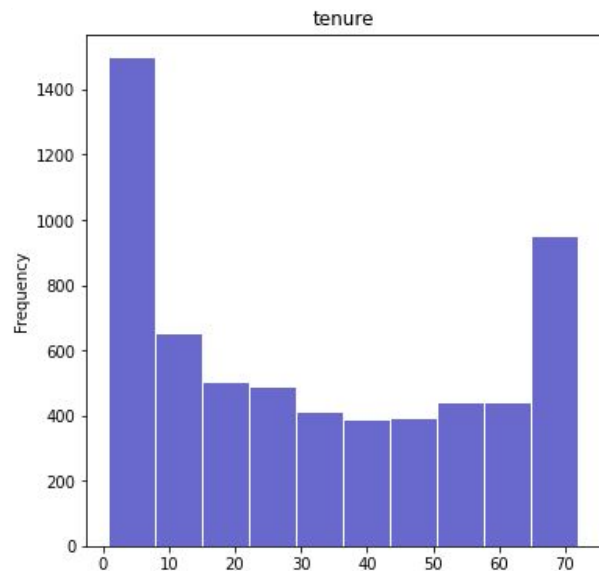


Categorical variables (3/3)

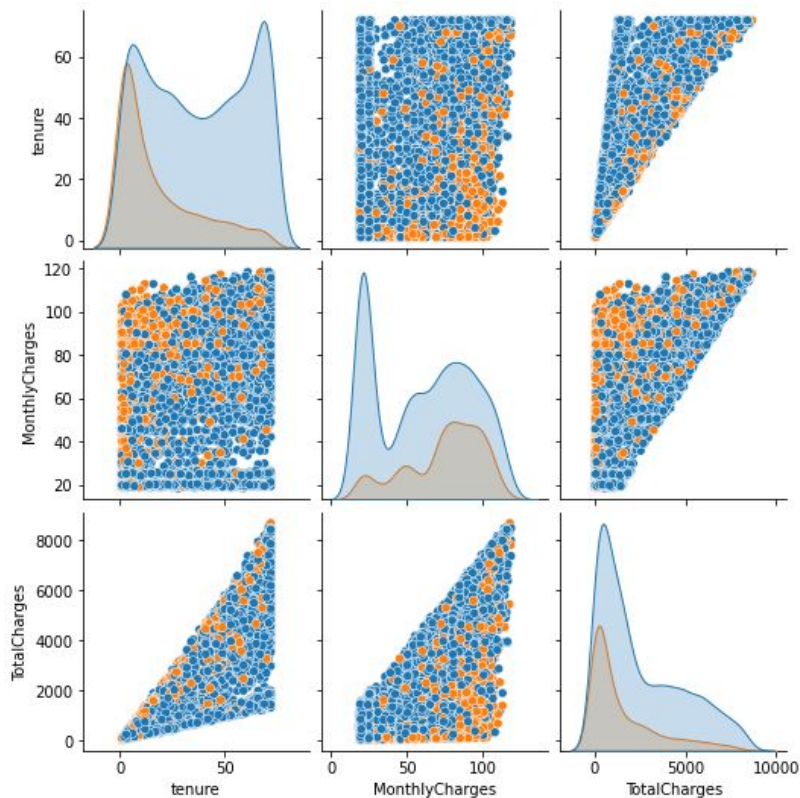
OnlineSecurity



Numeric variables



Numeric variables



Users with **lower tenure** and **middle-higher monthly charges** tends **to churn**.

Users with **higher tenure** having more total charges over time. Thus, the relation is linear.

Modelling

ML-models

Supervised learning

We'll use 5 models for the classification problem:

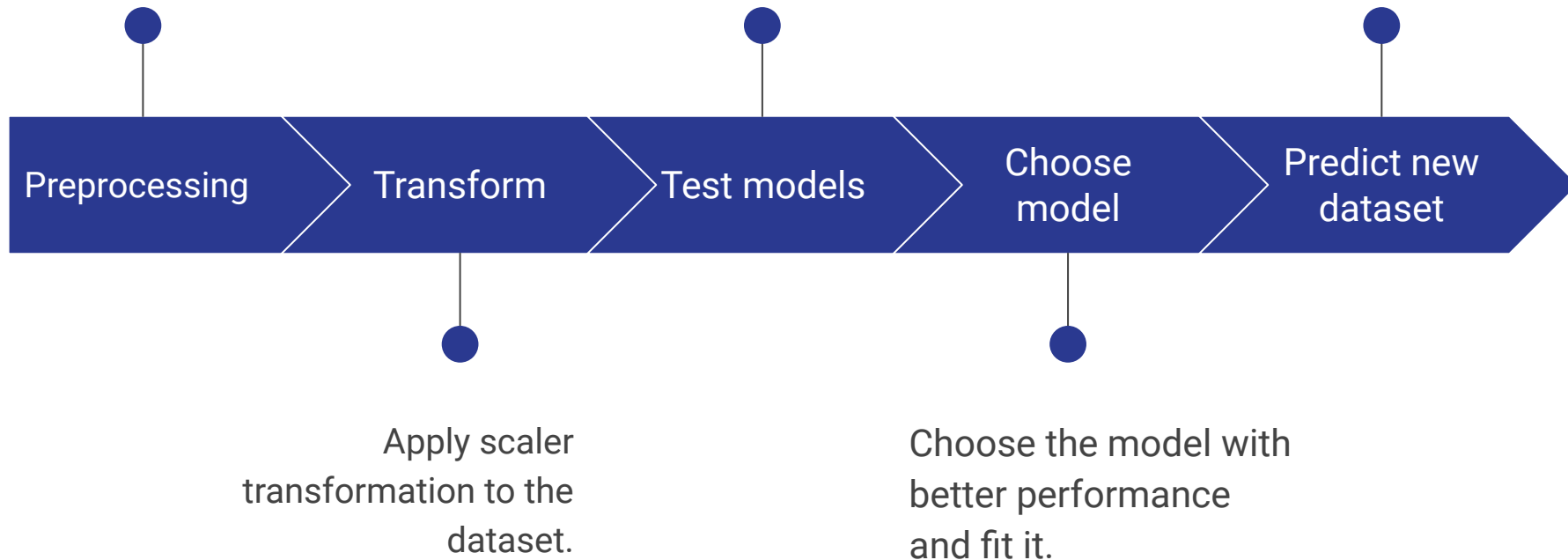
- Logistic regression
 - Gradient Boosting
 - Random Forest
 - SVM
 - K-Neighbors
-

- Split dataset in features and scope.
- Make dummies variables

From a list of models, test each one printing:

- Accuracy score
- Confusion matrix

Predict the churn behaviour of the new dataset and make recommendations.



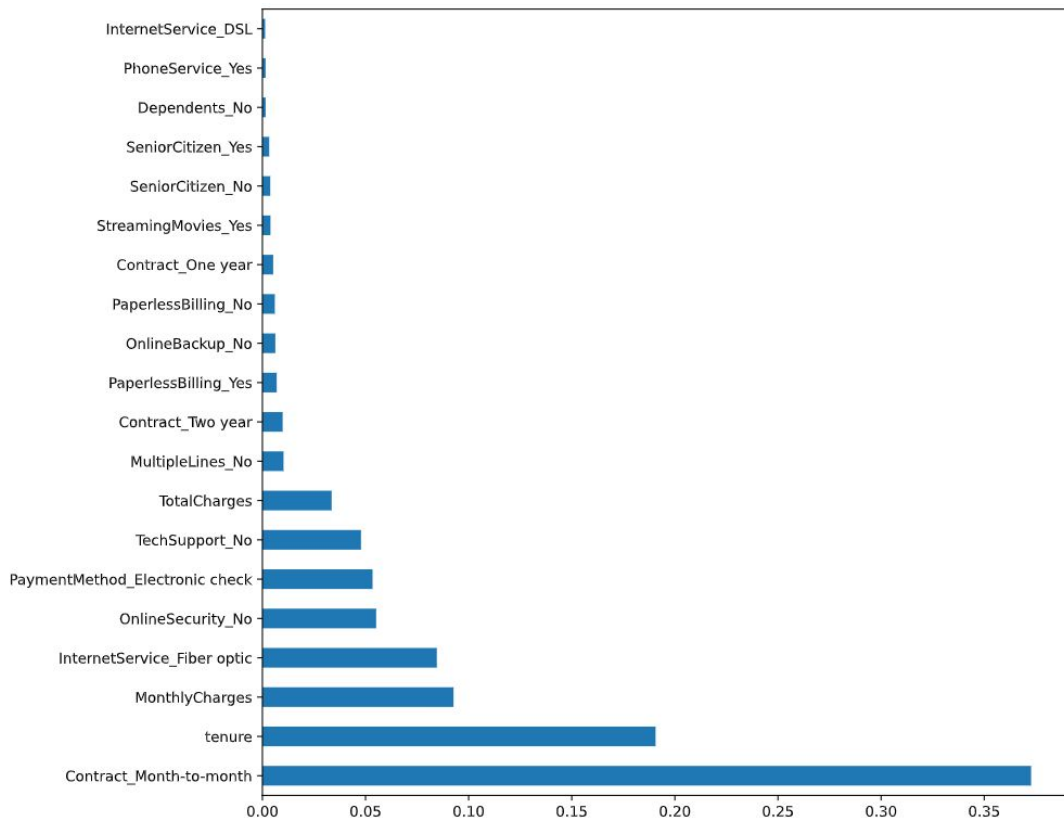
Accuracy score

```
Logistic Reg. : 0.785
Grad. Boosting : 0.803
Rand. Forest : 0.796
SVM : 0.782
KNeig : 0.762
```

Gradient Boosting shown the best prediction, we'll choose it.

Grad. Boosting				
	precision	recall	f1-score	support
0	0.84	0.91	0.87	4535
1	0.66	0.51	0.57	1597
accuracy			0.80	6132
macro avg	0.75	0.71	0.72	6132
weighted avg	0.79	0.80	0.79	6132

Variables relevance

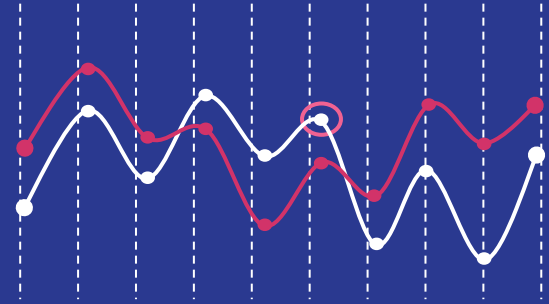


How we seen in the EDA, the **more important features** for predict the customer churn are:

- Contract: **Month-to-Month**
- Tenure
- Monthly charges
- Internet Service: **Fiber optic**
- Online Security: **No**
- Payment method: **Elec. check**
- Tech support: **No**

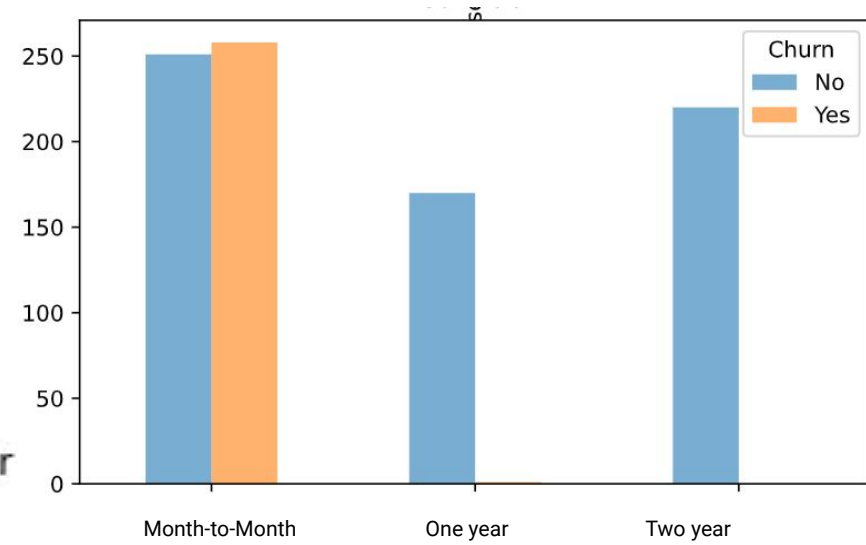
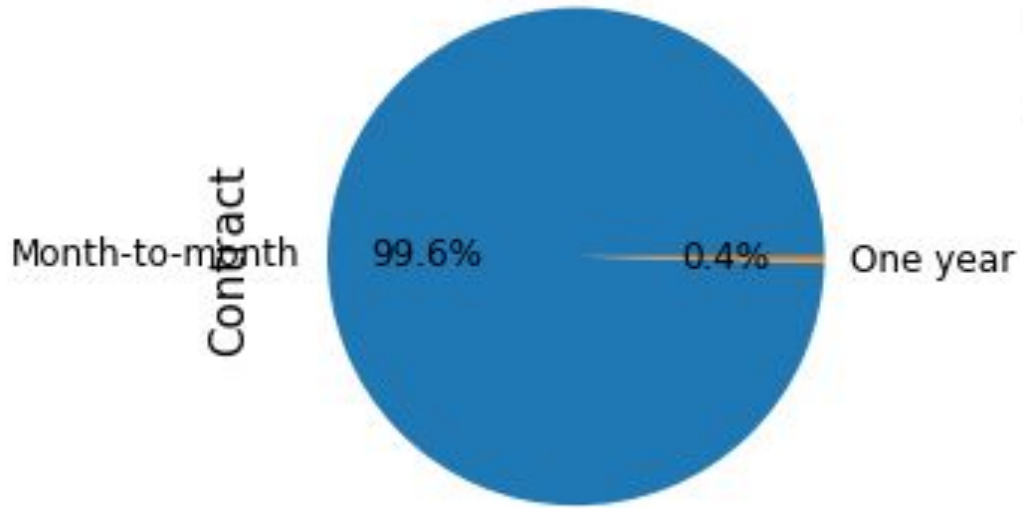
Predict

New database with possibility
users to churn service

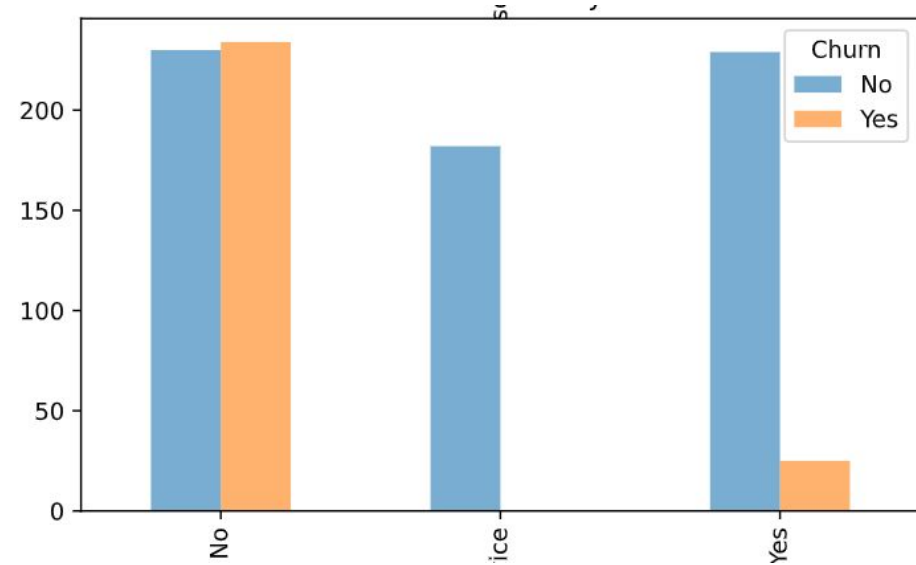
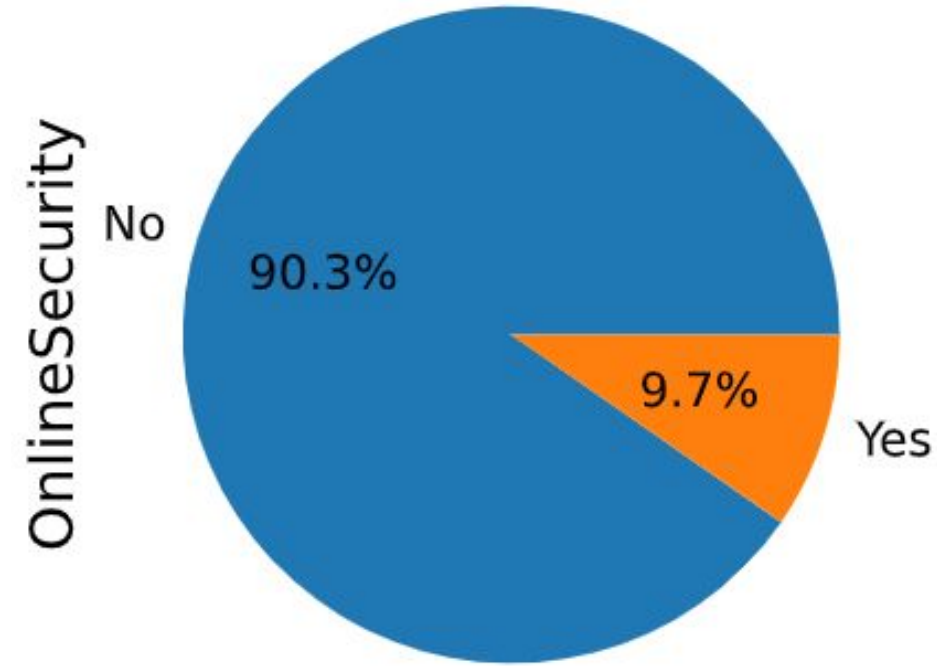


There are **259 users** with
possibility **to churn** the
services.

Contract



Online security



If we lose those
clients, we'll
lose more than
USD 19,000
next month.

Next steps

Improve the model performance

- Tuning the hyperparameters of ML models.
- Trying with more advanced models as Neural Networks
- Analyzing the bias and variance in the model performance.
